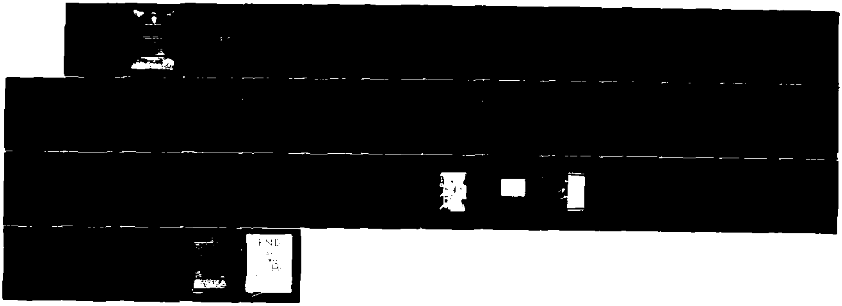
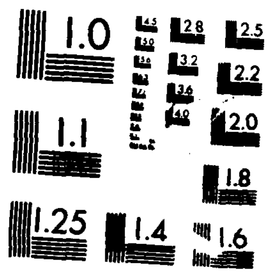
AD-A187 074 A FLIGHT EVALUATION OF VOICE INTERACTION AS A COMPONENT OF AN INTEGRATED..(U) ROYAL AIRCRAFT ESTABLISHMENT FARNBOROUGH (ENGLAND) R LITTLE ET AL. APR 86 1/1

UNCLASSIFIED RAE-TM-FS(B)-637 DRIC-BR-100581 F/G 1/3.1 NL



The next stage was to hold off the voice echo until the DVI activate switch was released. The echo was heard within half a second of release of the switch. This was a much more satisfactory method of implementation which was much preferred by pilots. In addition, the list of words which were echoed was increased to include the keywords such


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

(1)

ROYAL AIRCRAFT ESTABLISHMENT

Technical Memorandum FS(B) 637

Received for printing 22 April 1986

A FLIGHT EVALUATION OF VOICE INTERACTION AS A COMPONENT OF AN INTEGRATED HELICOPTER AVIONICS SYSTEM

by

R. Little

Lt Cdr R. Cowan RN

SUMMARY

A Wessex helicopter at RAE Bedford was used to develop and evaluate an integrated avionics system which incorporated advanced displays and a flight management system for both military and civil applications. Two important features of the system were automatic speech recognition and synthetic speech output. Flight trials have been conducted to establish guidelines for the successful integration of these devices with advanced avionics such as colour displays, digital maps and touch overlays. The use of speech technology in the cockpit offers an element of redundancy and if correctly integrated will be capable of improving the man machine interface to a far greater degree than is achievable by hand or voice alone. The trial has shown that data input and retrieval from such a well structured cockpit management system can be achieved quickly, simply and easily.

LIST OF CONTENTS

		Page
1	INTRODUCTION	3
2	THE SPEECH TECHNOLOGY EQUIPMENT	3
	2.1 The automatic speech recogniser	3
	2.2 The synthetic speech system	4
3	THE TRIALS PROGRAMME	4
	3.1 Speech recognition	4
	3.2 Speech output	5
4	SPEECH RECOGNITION RESULTS	5
	4.1 System reliability and packaging	5
	4.2 Recognition performance	6
	4.2.1 Vocabulary	6
	4.2.2 Active words/syntax	6
	4.2.3 Signal/noise effects	7
	4.2.4 Pilot variability	8
	4.2.5 Measured results	8
	4.3 Piloting aspects	11
	4.3.1 Speech quality	11
	4.3.2 Visual cues	11
	4.3.3 Voice echo	11
	4.3.4 Uses	12
	4.4 Integration aspects	13
5	SPEECH OUTPUT RESULTS	14
	5.1 Speech intelligibility	15
	5.2 Speech output control	15
	5.3 Operational effectiveness	16
	5.3.1 Priority information	16
	5.3.2 Optional information	16
	5.4 Voice interaction	17
6	CONCLUSIONS	18
Appendix A	The Wessex facility	21
Appendix B	Speech recognition and its implementation in the Wessex	23
Appendix C	Trials programme and procedures	26
Table 1	Speech recognition vocabulary	28
Table 2	Speech recognition syntax	29
Table 3	Total speech output vocabulary	30
Table 4	Priority speech output vocabulary	30
Table 5	Optional speech output vocabulary	31
Reference		32
Illustrations		Figures 1-10
Report documentation page		inside back cover

TM FS(B) 657

1 INTRODUCTION

A Wessex 2 helicopter at RAE Bedford was used to evaluate a suite of experimental avionics in flight. The new equipment comprised a set of colour CRT displays (one with a touch sensing overlay), automatic speech recognition (ASR), synthetic speech output (SSO), a monochrome control and display unit (CDU) (also with a touch sensing overlay), microprocessors for interface and algorithms related to engines, transmission, radios, guidance, airdata and navigation systems (including a digital map). The total cockpit was night vision goggle (NVG) compatible. The equipment and software were integrated to provide a user friendly advanced cockpit avionics and flight management system which could form the core of any future helicopter cockpit package. A detailed description of the Wessex facility is given in Appendix A and Ref 1.

The purpose of the development programme was two-fold:

(a) To determine how the total pilot cockpit display interfaces and supporting avionics could best be integrated and flight management data presented to a pilot, to optimise his performance, reduce workload and hence increase the overall mission effectiveness.

(b) To demonstrate a representative FLYING system to Service and Civil operators, and UK Industry and to stimulate their appreciation of the benefits that new technology offers when used to advantage.

This Memorandum, which is one in a series concerned with the sub-systems in the Wessex, describes the performance and operational use of the speech recognition and synthesis systems in the cockpit. These recent innovations have attracted much interest, in that it was hoped that their use would alleviate some of the high pilot workload associated with the airborne management of a sophisticated suite of avionics. The reasoning was that the use of speech technology, would enable the pilot to concentrate on flying, with eyes out of the cockpit. This would enhance flight safety.

During the trials the main thrust of the work was towards the integration of the new technologies, to discover when and how they should be used in combination with other technologies in the cockpit. Although the individual performance of the speech recogniser and synthesiser was measured, this was of secondary importance for two reasons. Firstly, the equipment used did not represent the standard which could be expected to be in Service in two or three years time. Secondly, criteria for sensible performance measurement of speech recognition and synthesis systems have yet to be defined.

2 THE SPEECH TECHNOLOGY EQUIPMENT

2.1 The automatic speech recogniser

The device used in the Wessex was the SR-128 speaker dependent word recogniser. It could accept phrases up to 8 s in duration and had sufficient memory to be able to recognise up to 128 s of speech utterances. When using the system the speaker had to provide an isolated utterance of each work in the vocabulary to be employed. The pilot initiated a command to the recogniser by using an activate switch on the cyclic flying

control. A full description of the SR-128 and its installation - the Wessex is given in Appendix B. Table 1 lists the vocabulary used in the Wessex.

2.2 The synthetic speech system

A synthetic speech facility which could be used to give audio messages or warnings to the pilot was designed and installed in the processing and interface unit. This was based on digitally recorded speech using a Texas Instruments 5220 linear predictive coding speech processor, which could store up to 500 words of vocabulary. Simple commands from the main processor could then select a word or group of words, its volume, and its priority for output if more than one word or phrase was called. Any vocabulary could be programmed by any voice, using a complementary portable speech laboratory specifically produced by Texas Instruments for that task. This facility enabled a user to program any words or phrases into programmable read only memory (PROM) which was then inserted into the speech interface card. The voice output was fed into the aircraft intercom system via an onboard amplifier. The synthetic speech could be totally isolated from the intercom by deselecting it on the intercom/radio station box.

Three groups of vocabulary were distinguished by the main processor system. The first corresponded to warnings such as engine failure. The second corresponded to optional information such as height calls used for the closing stages of instrument approaches. This level could be inhibited via the CDU, to eliminate nuisance calls when hover taxying or flying nap of the earth.

The third group corresponded to a voice echo of the words recognised by the SR-128.

Table 3 shows the vocabulary that was stored on the speech card.

3 THE TRIALS PROGRAMME

The man-machine interface of the Wessex cockpit was considered particularly important. The speech recogniser and synthetic speech output systems were on integral part of this interface. Speech recognition could be used in parallel with tactile controls such as keys and touch sensing overlays. Speech output could be used in parallel with the monochrome and colour displays. A description of the overall Wessex trials programme is given in Appendix C.

3.1 Speech recognition

To study the performance of the DVI system in flight many sorties were devoted entirely to its use. In this way not only was the speech recogniser performance measured but also the problems and attributes of using DVI during all modes of flight for a wide variety of tasks were assessed. On other occasions, pilots were briefed to use the DVI only when they wanted. They were free to use other methods of control such as switches or touch overlays at the same time. In this way it was hoped to discover the relative merits of different control systems and where each was preferred.

The speech recogniser could control the following functions

(a) Multi function display formats

- call digital map with tactical overlay
- alter map scale

- call engines display
- call hover display
- call check lists eg take off, shutdown etc
- call waypoint data such as ranges, bearings, ETA'c
- call flight status data such as endurance, all up weight
- call radio com plans
- call limits data (eg icing)

(b) Guidance

- select ILS or MLS sensors
- input runway headings
- input glideslope for MLS
- input cruise speed and height for the flight directors
- input decision height
- select go-around

(c) Communications

- select relevant radio
- select a preset frequency
- store a preset frequency
- input a radio frequency
- select a frequency by title (eg 'Bedford Approach')

(d) Navigation

- fix to a planned waypoint
- select a reporting base
- select a waypoint to put steering data on the PFD
- set and allocate a waypoint number to a chosen point on the nav plot
- update waypoint grid locations
- select or alter a route

3.2 Speech output

The speech output was used on all flights. The vocabulary was split into priority warnings and advisory information as already described. Tables 4 and 5 show the two groups of phrases and the events which the program used to prompt the speech output.

In addition the system was used to echo some of the DVI commands to give the pilot audio feedback on the word or phrase recognition that had taken place.

4 SPEECH RECOGNITION RESULTS

The evaluation of ASR can be split into several areas corresponding to system reliability, recognition performance, pilot use and acceptance, and integration aspects.

4.1 System reliability and packaging

The SR-128 equipment was powered by 240 v/50 Hz supplies and was designed for ground based applications. Luckily, it fitted into 19 inch racking and the internal power supplies could also be driven from the 115 v/400 Hz aircraft systems. The Wessex

was already fitted with 19 inch racks so the installation in the aircraft was straightforward. In the subsequent flight trials the equipment has proved to be very reliable failing only once in over 150 hours of test flying.

No interference effects were detected between the SR-128 and other equipment.

It is thought that with the present advances in VLSI technology and suitable power supplies the ASR equipment of the near future will be supported on a single card system which could then be an interface card in a processing and interface unit. This offers obvious benefits in system integration.

4.2 Recognition performance

The recognition performance of an ASR system has been found to be dependent on many factors. The figure of merit which is attached to this type of equipment is consequently without meaning unless other parameters concerning the ASR and the experiment which led to the results are known.

4.2.1 Vocabulary

The content of the vocabulary could have a great effect on the recognition performance. For example, most recognisers would perform well when distinguishing between the words 'speech' and 'voice', however the same recogniser would probably make more errors if the words were 'speech' and 'teach'.

This problem arose on many occasions during the initial DVI experiments in the Wessex helicopter. A good example occurred during navigational fixes over pre-designated waypoints. Unfortunately, the words 'fix' and 'six' were often confused resulting in poor recognition performance and pilot acceptability. To overcome the problem the word 'plot' was adopted in place of 'fix'. Pilots quickly became accustomed to the word and recognition performance increased considerably. Similarly 'ok' was preferred to 'enter' as the executive command. The vocabulary is shown in Table 1.

4.2.2 Active words/syntax

As the number of words which the ASR system was required to recognise increased the recognition performance degraded. Although, ideally, all the words in the vocabulary would be active at all times it was found that syntax was required to reduce the number of active words at any one time and thus preserve recognition accuracy and improve recognition speed. However, syntax introduced its own problems if it was not properly structured.

As with complex menu structures used in modern Control and Display Units, if the syntax was too restrictive pilots found the system unfriendly. This was particularly apparent when the ASR mis-recognised the pilots words and diverted the software to another area (called a node) of the syntax tree. This could result in the pilot thinking he had manipulated the programme correctly when in fact he had not.

A delicate balance between recognition performance and 'user friendliness' resulted in a syntax structure which was very similar to the Wessex's tactile system, based on the CDU menu and dedicated keys. This was not entirely surprising as great

effort had been taken to keep the CDU control 'user friendly'. The maximum number of words which were active at one time was thus reduced from 66 to 33. The major advantage of such a system was that the pilot used the same syntax whether he used the DVI or the touch overlay and keys. In both cases the CDU display reacted in the same way and could be used to remind the pilot of the active words available at any time. This of course required the CDU to follow the ASR programme and vice-versa. Thus commands started on the CDU could be finished by voice or the other way round.

4.2.3 Signal/noise effects

One of the greatest concerns at the outset of the programme was the capability of the ASR system to cope with the high background noise and variable signal levels in the helicopter. Measurements during flight indicated that the noise level in the cockpit was 105 dBA which was about 20 dB greater than the noise on the flight-deck of the BAC 1-11 which had successfully flown the SR-128 in earlier trials.

To reduce the effect of ambient noise the SR-128 employed a noise mask technique, see Appendix B. The technique was to use the SR-128 to sample the ambient noise, store it as a mask and use it to reduce the effect of noise when the ASR was used. The success of this technique was dependent on the power spectrum of the noise being stationary. Trials in a Buccaneer at RAE Farnborough and the BAC 1-11 at Bedford had experienced problems using the SR-128 due to the dependence of the noise on aircraft speed, power setting and configuration. To retain adequate performance repeated noise samples were required as the flight regime changed. This was time consuming and tedious. Fortunately, the problem was not so severe in the Wessex, since the principal noise characteristics were dominated by rotor and transmission systems which rotate at governed rates. Power changes in helicopter turbine engines do not produce large variations in engine compressor noise when compared to the background transmission noise. As a result one sample of noise sufficed for most conditions during a flight. Having said that, there were several occasions when recognition deteriorated for no apparent reason. When this occurred noise masks were taken on the ground, in the hover and in forward flight and this generally improved the situation somewhat. Fig 4 shows the mean noise mask levels in each of the ASR channels for different flight conditions. On another occasion when recognition was particularly poor, changing the pilots microphone and retaining the voice template dramatically improved recognition from about 50% to nearly 100%. This highlights a major area of concern for the introduction of a DVI system into operational use. Existing equipment such as microphones will have to be considered as part of the DVI and as such may need more stringent tests than are currently used to check for adequate intercom performance. It also highlights the need to ensure that the pilot's microphone was positioned correctly and, moreover, that it had the same characteristics as when it was originally used to prepare the voice templates.

Other factors which were found to affect recognition were the volume of the radio and the intercom. Loud radio tended to make pilots speak more loudly and paradoxically loud intercom made them speak more softly. The effect of both of these not only varied

the signal to noise ratio but also the speech characteristics of the pilot. Both effects reduced the recognition performance.

As the templates had been created at a fairly constant mean volume any variations from this in operation could degrade the pattern matching technique. The SR-128 appeared to be particularly sensitive to volume variations which resulted in the pilots having to talk to the DVI in a deliberate unnatural manner in order to obtain adequate recognition performance.

To overcome these effects future ASR systems must employ active noise compensation techniques. Furthermore a system which employed automatic gain control (AGC) of the microphone level would reduce the variability of the signal/noise ratio, although the pilots speech characteristics would still change with varying loudness.

4.2.4 Pilot variability

During some of the RAE trials the word error rate was found to vary between pilots. The variations were not random. Pilots who achieved good recognition performance did so regularly. Some pilots however, produced inconsistent results which were invariably worse than the results obtained from the 'consistent' pilots. The variation was as much as 5:1 in error rate from typically 2-10%. A technique which has been evolved at RSRE Malvern was used to determine if this problem was a function of the ASR or the pilots themselves. Pilots were asked to record a number of words and phrases many times. By correlating utterances of the same word it was demonstrated that some pilots, even when trying to speak in a repeatable manner, produced a large spread of acoustic patterns compared to other pilots. Consequently, any ASR which used acoustic pattern matching techniques would produce results which ware dependent on the individual user. The significance of this problem has yet to be established, particularly the variations that may occur over a wider population of pilots with different nationalities and accents.

Future ASR systems will need to cope with within-speaker variability as pilot selection on the basis of DVI recognition performance is not likely to be favoured by operators or pilots.

One possible solution being investigated by a number of speech technology groups involves Hidden Markov Modelling techniques to model speech variations. It is subtle changes due to stress, health, age etc that would need to be tackled by such a system.

4.2.5 Measured results

The results presented in this section relate to flights when pilots were briefed to use the DVI system only when they wished to do so because it was considered the most attractive option under the circumstances prevailing at the time. They were free to use other methods of control such as the keyboard or touch overlays when they felt it would be easier. All of the pilots used a boom microphone. It was found that the use of a foam sock over the microphone significantly reduced breath noise. Although this did not appear to effect the SR-128 recognition performance, audio recordings were much clearer.

Measurements from initial pilot/DVI training flights and flights during which equipment problems were uncovered have not been included in the analysis. The results are therefore representative of three trained pilots using a fully operational system.

(a) Overall recognition performance

The total number of words or phrases which were used in the analysis was 5290 of which 345 were mis-recognised. This corresponds to a recognition rate of 93.5%. The three pilots used on the Wessex trials recorded performances to within 1% of each other. It is interesting to compare this with the results from the BAC 1-11 trials in which consistent differences in performance were obtained between the trials pilots. The similar performances in the Wessex may be due to the high ambient noise condition which prevailed in the helicopter which was about 20 dB greater than that in the BAC 1-11. This high ambient noise may have induced the pilot's to articulate more precisely to combat the noise and prevented them speaking as naturally as they would on the flight deck of a jet transport aircraft. Thus the performance of the SR-128 recogniser in relatively noisy environments may be less dependent on the individual characteristics of the pilots.

(b) Word or phrase recognition performance

The recognition performance of individual words or phrases is shown in Figs 5 and 6. Recognition rates for individual phrases vary between 53% and 100%. Although detailed statistics of individual words in the DVI vocabularly are unreliable, given the small number of samples, some important features can still be observed.

The phrase 'tees and pees' was successfully recognised on only 53% of occasions. If this word is removed from the overall performance analysis then the recognition rate increases from 93.5% to 94.5%. It is interesting to note that this word was used to call up an engine monitor format on the colour MFD. If the phrase was mis-recognised then the pilot had simply to repeat the phrase. Although pilots found such frequent errors very annoying they considered the benefits of speech sufficiently great to persist with its use.

The effect of small changes in the vocabulary can be seen by comparing the two phrases 'map up' and 'map down'. These two phrases were used in identical areas of the syntax and were used by the pilot either to display the navigation format on the colour MFD or to change the range scales (hence up or down) if the format was already displayed. However, a 7% difference in recognition rate for these two phrases has been measured.

The executive command 'ok' was the word used most during the trials. It recorded a recognition rate of over 97%. The executive word for 'clear' which was 'rubout' achieved a 100% recognition rate. Low error rates for executive commands such as these are obviously important. It should be remembered that the commands 'enter', 'clear', 'cancel', 'confirm' were all tried but none were so successful as 'ok' and 'rubout'.

The main goal of the programme has been to establish what to do with DVI rather than to find out how good current recognisers are. To achieve this, the vocabulary was tailored to include unique words which were unlikely to be required in future recognisers. Carefree speech cannot be handled by speech recognisers yet. As a result, strange words have appeared in the CDU programme such as plot (fix), fife (five), niner (nine), ok (enter), rubout (erase/clear). A major benefit of better recognisers would be the ability to choose a vocabulary to suite the pilot instead of the recogniser.

Because aviation voice procedure is well established, changing the associated vocabulary to suit speech recognisers would introduce its own problems. Therefore the trials vocabulary is seen only as a temporary measure.

(c) Digit recognition

The overall recognition rate for the digits 0–9 and the executive ok and rubout commands was 95.4%.

A confusion matrix for the digits is shown in Fig 7. The recognition performance varied between 99.4% for the digit '3' to 90% for the digit '4'. It is also interesting to note that the digits '7' and '8' were never confused with other digits.

When quoting recognition error rates for connected or continuous word recognisers one has to be extremely cautious. Although a 'word' error rate is of great interest and use when comparing the performance of different equipments, care has to be taken when using the results as a measure of operational effectiveness. Consider for example, a defined word error rate of 5%, ie a 1 in 20 recognition failure rate. If a pilot now uses this system to input a series of 10 digit strings for navigation coordinates then a 50% navigation update error rate will result which is not so attractive. Of course, the rate with which a pilot can input data is also important and it should be pointed out that a typical input error rate for a keyboard is of the order 5%.

(d) Word usage

Figs 8 and 9 show the DVI word or phrase usage rate for the three pilots. As can be seen all three pilots used the DVI vocabulary in similar relative ways. Data entry was the most used feature of the DVI. In addition, selection of key areas in the CDU structure such as comms, navigation, guidance or menu was accomplished many times by voice. Display mode selections, which could be done at any time by one key depression were also made by voice, for example, map up, map down and tees and pees.

There are clearly defined situations when speech recognition is the preferred method of control or data insertion and others when equally clearly it is not. Between these two extremes the relative advantages of voice and tactile control are less clear and the key to success lies in the integration of the two.

In a noisy radio environment, with a busy navigation task, plus perhaps a need to set up the guidance page on the CDU for an approach, DVI is not easy to use. This is particularly true if there is also a high level of intercom chatter. The initiation and maintenance of any conversation absorbs the pilots time and concentration. The busy pilot may be listening to two radios plus the intercom, and also talking to one agency by radio and other aircrew on intercom. He hardly wants to open a rigidly structured conversation with his aircraft too. Nevertheless, the pilot may choose to use voice for parts of his cockpit management, even when busy. In these situations the recognition performance must be very good.

4.3 Piloting aspects

4.3.1 Speech quality

The pronunciation of many words in normal speech is dramatically affected by its context. For this reason when the system was trained, it was necessary for the pilot either to anticipate the application context or else say the vocabulary in a neutral manner and without emphasis. Pilots found they frequently had consciously to attempt to recreate the word as trained. The use of speech recognition is clearly a skill which has to be acquired.

Pilots using DVI for the first time often needed three or four flights, with interim re-training periods, before the recognition performance of the DVI was at the high levels described in section 4.2. This was a function primarily of the time needed to acquire this 'DVI skill'.

4.3.2 Visual cues

As errors can occur in word recognisers, it is necessary to feed back to the pilot the results of the utterance. One direct method used in the Wessex was a read out line on the CDU monochrome display to tell the pilot what the ASR had recognised see Fig 2. Initially this line was invariably checked but less so once voice echo was introduced.

It was intended during the research programme that the DVI programme and its syntax tree should closely resemble the CDU page structures. This meant that the CDU prompted the available DVI vocabulary thus eliminating the possibility of the subject pilot finding himself in a corner of the programme wondering how to get out.

A problem which was noted early on in the trials was the inability of the ASR syntax to track that of the CDU although the CDU was able to track the ASR syntax. Thus if the ASR was left in one mode whilst the CDU moved to another and if the ASR was not then moved to the new node, chaos could ensue, ie the system was unfriendly.

The difficulty has now largely been resolved although the DVI takes a second or two to follow the CDU. This caused difficulty at times when mixing input channels quickly. Essentially in a system to be introduced into service mutual tracking must be fast and transparent to the pilot.

4.3.3 Voice echo

Although the read out line on the CDU was very effective there were situations during which the pilot could not afford the time to scan the display. The use of the synthetic voice system to provide a voice echo of the recognised word or phrase was therefore investigated.

Initially, the digits 0-9 and the executive command 'ok' were echoed. The echo was initiated as soon as the recogniser output its data string to the microprocessor which hosted the synthetic voice card. This frequently resulted in the voice echo being initiated while the pilot was still using the DVI. This often upset the rhythm of the pilots DVI commands. Nevertheless, pilots found the audio confirmation useful and they found themselves scrutinising the read out line and the CDU data entry fields less often.

The next stage was to hold off the voice echo until the DVI activate switch was released. The echo was heard within half a second of release of the switch. This was a much more satisfactory method of implementation which was much preferred by pilots. In addition, the list of words which were echoed was increased to include the keywords such as 'comms' and 'navigation', and also some of the more commonly used words such as 'plot'. Many of the DVI commands related to display mode changes were not echoed as this would unnecessarily create excessive audio. If a request for a display mode had been made then the pilot would be looking at the display and there would be no need for audio confirmation.

Frequently mis-recognition patterns were dominated by the same word. This would be no problem if the word was 'Machinhamish Talkdown' but if it was 'two' then the situation would be more serious. Increasing the number of syllables clearly improved recognition performance but a significant number of utterances, such as the digits, were constrained to remain monosyllabic.

Digits caused major problems initially but with pilot experience and an improvement to the SR-128, this difficulty was largely overcome. The ASR may well recognise 95 out of 100 different utterances. When inputting large strings of digits or commands the individual word error rate becomes less important. The important factor is the probability of inputting a complete communication sequence. Word error rates of 5% may then become communication sequence error rates of 50% which is not satisfactory.

4.3.4 Uses

The major benefits of DVI were not immediately obvious since they generally occurred as a result of integration.

Reading in strings of digits such as 10 figure groups for waypoints was particularly useful and much quicker than the manual equivalent. When inputting digit strings by hand from a list the pilot has to read the list, look at the keyboard whilst keying, and then the CDU to confirm accurate entry. In addition people frequently spoke the digit string whilst keying. However, the action can be reduced to reading the digit string out loud and checking recognition on the CDU screen or listening to the voice echo. This aspect of integrated DVI and SSO was found particularly useful. Another frequent use of DVI was to change the MFD formats particularly in hovering or NOE flight, with the pilot flying with hands on controls at all times and eyes predominantly outside the cockpit.

At other times combinations of hand and voice were used to input data.

During periods of intense concentration such as in unstabilised instrument flight some pilots tended to use their hand to select the display element required and then perform the data input and execution command by voice, thus the keyboard search task was eliminated.

In busy audio environments pilots tended to favour manual input to speech input. This was particularly true of simple selections where the pilot could act instinctively.

Using the DVI when listening to a radio message was difficult. The pilot had to concentrate before beginning the conversation with the cockpit and this could easily lead to him missing a radio or intercom message.

An interesting problem was encountered when one pilot returned from a German visit with a voice template shift. The pilot had become accustomed to using different speech inflexions to be understood by German Air Traffic Control. A speech template trained immediately after he returned was quite useless within two days as his speech pattern returned to normal. Retraining the template cured the recognition difficulty. This was an interesting phenomenon which could have ramifications when flying from one country to another where the pilot needs to apply an inflection to his voice in order to be understood by air traffic.

4.4 Integration aspects

From these trials it is clear that DVI will often be used as the primary method of control of systems and of data entry. However, because of the operational problems which can sometimes arise it is unsuitable as the sole means of controlling the aircraft systems. Voice recognition should therefore be regarded as an additional control mechanism which can be used as an alternative to tactile methods of control such as touch overlays or keyboards. This will provide a useful level of redundancy should one of the tactile input systems fail, but more importantly speech provides a valuable alternative whenever tactile control is inconvenient for the pilot to use.

A DVI system, if properly integrated with other cockpit systems has been shown to be extremely effective. Without careful integration the DVI system will prove 'unfriendly' and in the limit unusable by pilots. The trials have established guidelines for integration in three areas

(a) DVI enable

The DVI enable mechanism which was used in the Wessex, relied on the pilot using a flight control mounted switch. This switch relayed the pilot's microphone to the DVI. The pilots found this position easy to use and they could fly 'hands-on' throughout. Subsequent DVI messages were limited only by the syntax needed to maintain adequate recognition performance and sensible correlation with the CDU programme. The technique adopted was identical to that when using the radio 'press to transmit button', and as such pilots found it a natural control.

A method of enabling the DVI which was considered but not adopted was that of using distinctive keywords to engage and disengage the DVI, over a 'hot microphone'.

Pilots disliked the keyword philosophy for a number of reasons

(i) The keywords would have to be carefully chosen so that they were distinctive and rare in normal conversation. Although accidental engagement might be rare, nuisance DVI activation would be very annoying.

(ii) Keywords would increase the transaction time, particularly for simple one word commands.

(iii) The pilot would have to remember to use the de-activate keyword at the end of every communication with the recogniser.

(iv) Pilots disliked any form of artificial limitation to the use of DVI. In particular, rigid syntax was criticised. Keywords represent a form of rigid syntax which when combined with other syntax within the recogniser programme would require an unnatural form of pilot-DVI communication.

(b) Tracking of I/O systems

With a multitude of methods of controlling systems and inputting data, it was found to be imperative that each control system could track the operations of others in order that any type of control could be used by the pilot at any time. For example, the syntax used within the DVI must track the CDU menu structure and vice versa. If this was not done pilot workload could increase significantly when an operation started with one input channel could not be completed by another.

(c) Feedback of the word recognition process

When using DVI, it was important for the pilot to know what word or phrase the ASR actually recognised. Often this was obvious, for example the correct multi-function display format, such as the map, would appear when requested. However, on those occasions when the system did not react or mis-recognised a command, the pilot found that feedback of DVI recognition was essential. Two methods of providing this feedback were investigated.

The first was a visual read-out of the previous and current outputs from the ASR, which was displayed as a feedback line on the monochrome display of the CDU (see Fig 2). This was very successful enabling a pilot quickly to identify and correct errors caused by mis-recognition or rejections (ie say it again).

The second was a voice echo using the synthetic speech system. The main benefits of voice echo were noted during the insertion of long strings of figures such as way-points and frequencies in low level hovering and NOE flight. Pilots found their ability to lookout was much improved by voice echo since there was no need to look in and focus on the CDU to check data input.

The voice echo was also useful when errors occurred. Pilots quickly picked up wrong or absent responses, such as a wrong number or a recognition error which took the programme to the wrong node of the syntax.

5 SPEECH OUTPUT RESULTS

The synthetic speech system was used during most flights so that its effectiveness could be evaluated in a variety of flight regimes. Three types of output have been investigated corresponding to priority information, optional information, and an audio feedback of the DVI recognition process as the first stage of a voice interactive system study.

TM FS(B) 637

5.1 Speech intelligibility

The vocabulary required for the trials could be produced using a portable speech processor system. Prior to the flight trials several people, male and female, from many occupational areas, were invited to program a sample vocabulary which was subsequently tested in the laboratory. This was not a sophisticated test but simply consisted of a few people comparing the resultant quality of speech. On this basis the voice of a female secretary was judged to be the most intelligible and distinctive.

During flight, the intelligibility was not as good as in the laboratory but was nevertheless satisfactory. The distinctive female voice could always be distinguished from other aural sources such as helicopter crew or air traffic controllers etc. In particular the word 'warning' which prefaced some of the messages was an excellent attention-getter which never failed to focus the pilots thoughts immediately to the problem that had emerged.

The elements of the synthetic voice used for echo purposes were not so intrusive that they blotted out incoming radio traffic or essential intercom chatter. Under such circumstances, pilots quickly became accustomed to checking the CDU and filtering out the synthetic speech.

The SSO in echo mode operated in a somewhat uneven staccato manner which was due to the speech processing system. To improve this the synthetic voice will need to be clearer, faster and delivery will have to be better measured with similar gaps between words. The system used could have been improved but considerable time and effort would have been required to edit the stored vocabulary. In the future a time synthesis-by-rule system, as opposed to one based on recorded speech, may be the best way of achieving acceptable prosody.

5.2 Speech output control

When the aircraft power supplies were activated prior to engine start the voice card was programmed to output a time dependent recognition phrase such as 'good morning'. In this way the pilot was assured that the voice facility was operational.

Certain words or phrases, listed in Table 4 remained active at all times. The majority of these related to failures or limit exceedences, which occurred infrequently, unless deliberately induced. The appropriate single warning was given, such as 'check fuel'. This was not repeated unless the fault disappeared, or was rectified, and then developed again. The torque margin output was repeated every 7 s if the limit continued to be exceeded.

The second group, listed in Table 5, corresponded to optional information which was all related to height and height clearances. All outputs were repeated every 7 s if the cause of the initial call still remained. To avoid nuisance warnings which could occur during some modes of flight, this set of calls could be disabled by the pilot using a control on the CDU display. When disabled in this way the voice system announced it had been disabled by saying 'goodbye'.

The group of calls could be enabled by selecting the same control on the CDU, this prompted a 'hello' to confirm the system was again operational.

This re-activation process also took place via the control program. When ILS or MLS were selected the SSO optional vocabulary was activated as the aircraft passed through 500 ft QFE (or rad alt) in the descent. Thus the SSO said 'hello' at 500 ft and then continued down to decision height in 50 ft steps.

It is worthy of note that even though pilots were working hard on some approaches they looked forward to the 'hello' call followed by the height count down despite the natural tendency to filter out extraneous audio information. This made these calls particularly beneficial. The height calls and pull up command were also used during low level night operations. When hovering or landing the calls were usually inhibited to avoid nuisance. To ensure that the system was re-activated if the pilot had forgotten to do so himself, the system was automatically reselected when the groundspeed increased through 30 kn.

5.3 <u>Operational effectiveness</u>

5.3.1 <u>*Priority information*</u>

Although under normal circumstances most of the warnings would occur infrequently all could be artificially induced by various means. Whenever a problem was discovered several visual warnings in addition to a voice call would be given to the pilot. For the most critical problems, such as an engine failure, the multi function display format was automatically presented with a format which showed the problem together with a small list of checks. In addition the CDU display was cleared except for a short message which repeated the voice call. The next level of warning, such as a fuel contents mismatch between tanks, produced the voice call and a message on the CDU. The final level produced only the voice call. It should be noted that all the warnings, whether a display format change occurred or not, produced colour changes on at least one of the formats. For example, if the torque limit was exceeded the torque strip on the PFD would change colour from white to magenta. In this way any speech output warning was reinforced by a change in format and/or symbology colour change. This combination was very effective and attracted the pilots attention to the salient feature of the display.

If the pilot was head out on a map of the earth exercise, a simulated cable warning could be programmed to appear during the sortie. On all occasions the speech output attracted the pilots attention, and on looking at the displays the problem was quickly diagnosed confirmed and dealt with. The torque limit warning often occurred when in low forward speed or hovering flight. On hearing the warning 'torque' pilots would often lower the lever, if it was safe to do so, without looking immediately into the cockpit, and thus concentration on the visual scene could be maintained.

5.3.2 <u>Optional information</u>

A list of the optional vocabulary is shown in Table 5. These words were usually enabled by the pilot (or the main processor) during instrument approaches or transits. The majority of the vocabulary was used in the generation of aircraft height calls

during an instrument approach. When the decision height was reached a 'decision height' call together with a DHT flag on the PFD would be activated.

A programmable low height warning could be selected by tne pilot. If the aircraft descended below this height a 'pull up pull up' message was generated. This was used during any flight condition when maintaining a set height clearance was essential.

This was particularly useful during any operation which demanded minimum terrain separation. The feature was specifically introduced for low level NVG operations. Height assessment during these operations was far from being easy or accurate due to the limited field of view of the NVG's and the monochrome, artificial nature of the image seen by the pilot. In addition to the pull up command the height calls at low level gave useful trend information. For example, when flying at 100 ft above ground level the pull up command could be set to 70 ft. Height calls at 100,90 and 80 ft would be heard before the final 'pull up pull up' command was given.

When the height director mode was selected on the PFD two further calls corresponding to 'high' and 'low' were enabled. These were used to define height boundaries about the height profile that was flown on the height director. They were reinforced with flags on the PFD.

Pilots liked the advisory information and warnings related to height. Although they did not reduce workload pilots felt more confident when they were present. There is little doubt that such a system would have a tremendous impact on the safety of both civil and military operations. The ability to disable the calls at the pilots discretion was considered vital to avoid unnecessary chatter at the end of an approach or during general visual flight. Without this facility the synthetic speech would be a nuisance. Pilots would switch it off and the benefit of safety calls would be lost too. System logic to ensure that the speech output was enabled prior to an approach has proven to be effective.

5.4 Voice interaction

This has already been discussed to some extent in section 4.3.3 which described the use of synthetic speech as a feedback to the pilot of DVI recognition. The success of a full voice interactive system will depend heavily on the operating environment in which it is asked to work. During the Wessex trials two comments were often made.

(a) There were times when DVI operation was difficult due to other audio tasks.

(b) The speech output should be kept to an absolute minimum to avoid it masking other radio and crew messages and becoming a distracting nuisance.

Having pointed out these limitations there may still be many occasions when question/answer sessions between a pilot and the avionics systems could be effective and worthwhile. Such a system would require complex computing but, with current technology, is a realistic proposition. The acceptability and effectiveness of soliciting synthetically spoken information by voice will be addressed in future trials.

6 CONCLUSIONS

The Wessex flight trials have successfully demonstrated the capabilities, benefits, and problems of using speech recognition and synthesis systems in conjunction with other advanced cockpit facilities. The technologies integrated with the speech systems included colour CRT displays, digital maps, touch overlays, joysticks, processing and interfacing to other airborne sensors and equipment such as engines, transmission, radios, guidance, navigation, and airdata systems.

(a) Speech recognition

(1) The performance of a speech recognition system has been found to be very dependent upon vocabulary content, syntax governing the number of active words, ambient noise and speaker characteristics.

(2) During trials in which pilots were briefed to use the DVI only when they felt it would be more useful than other control mechanisms, a mean word error rate of 6% was recorded. On some flights 100% success was achieved.

(3) Pilots found the speech recognition system extremely useful for many different tasks such as navigation updates, radio management and display mode selections. It was particularly useful when visual hand/eye co-ordination tasks were absorbing considerable pilot effort, such as during nap of the earth flight.

(4) Operational conditions have been encountered when the DVI system was difficult to use and other methods of control, such as the touch overlay, were favoured. These conditions were associated with a busy audio environment, for example during the marshalling phase of an approach. Under such circumstances DVI was difficult to use because of the random nature of incoming information and a need for the pilot to gather his thoughts before speaking.

(5) DVI systems need to be properly integrated within the core avionics system if the potential benefits of speech recognition are to be fully realised. In particular the following are required

- a method by which the pilot can activate the DVI using a cyclic or collective mounted switch, thus enabling hands on control.
- a core avionics system capable of controlling all input and output devices to enable the pilot to commence an input with one system and complete it with another.
- a feedback, both visual and aural of all or part of the recognition process to the pilot.

(6) Speech recognition systems need to use AGC on the pilots voice input microphone level or have recognition algorithms which are tolerant of the varying signal level of speech that occur naturally in everyday conversation and in the cockpit.

(7) Active noice compensation techniques are required to avoid the requirement to load 'noise masks' into the recogniser, a process which is both tedious and time consuming.

(8) To be operationally acceptable future recognition systems must be tolerant of the variable nature of pilots speech during all phases of flight as well as during stressful conditions.

(9) Random word recognition success rates in excess of 99% are required if DVI is to become a realistic tool in the helicopter cockpit of the future. This will provide an average communication sequence error rate of 5% which will be acceptable. Although systems are incapable of meeting this target at present, future systems show every indication of having sufficient performance.

(10) During the course of a helicopter mission the loading placed on the sensing and control mechanisms of a pilot vary considerably. At times, for example during low level flight, his hands will be fully committed to the flying controls and his eyes will be concentrating mainly on outside features. Under these circumstances DVI can be extremely beneficial. At other times, for example during complex cockpit crew procedures, in busy radio environments, or in high stress situations, the aural channels can be saturated or degraded. Then touch overlays and hands-on controls can be superior. Between these two extremes a blend of input methods is required.

For these reasons it is not envisaged that DVI will replace any control mechanism. It offers an element of redundancy and if correctly integrated into the cockpit it will be capable of improving the man machine interface to a far greater degree than hand or voice alone.

(11) The training of voice templates for the DVI system needs to be considered more fully in operational situations. Although time can be expended in a research environment to train the DVI carefully and then edit any anomolies in flight, this would not be cost-effective or realistic in operational use. Pilots will require a one pass guaranteed training session which should be possible to accomplish off the aircraft.

(12) The method of loading pre-recorded templates into the DVI system needs to considered. This should be a simple, quick process, which the pilot can initiate from the cockpit.

(13) Future DVI systems should allow a multi-crew operation. This does not require speaker independent systems but simply the capability of storing and using more than one set of voice templates.

(14) Limited syntax improved ASR performance but excessive use of syntax degraded user-friendliness of the total system. Future systems should seek to remove rigid syntax by the use of further post-recognition intelligence and dynamic control of some of the key recognition parameters such as distance thresholds and weighting factors.

(b) Speech output

(1) A speech output system based on linear predictive coding techniques and using a female voice had been proven to be intelligible and distinctive in the helicopter cockpit environment.

(2) The use of a 'warning' keyword before critical messages was found to be a very useful attention-getter, which pilots reacted to very quickly compared to other less-critical alerts.

(3) Two levels of speech output have been found to be required. The first corresponds to high priority information which occurs infrequently. These should not be selectable and for the majority of cases they should produce a single voice warning. The second level corresponds to height related calls and warnings. These can occur frequently, sometimes during operations for which they were not intended. These should be pilot selectable to avoid becoming a nuisance.

(4) The optional height call and warning system would improve the safety of both military and civil operations. Pilots felt more confident when the speech system was enabled during for example, an instrument recovery or during low level flight at night.

(5) The use of speech output to provide indications of failures allowed pilots to stay head out without constantly needing to monitor displays. The use of voice, display format changes, and colour symbology changes to highlight problems to the pilot was found to be a very effective combination.

(6) Speech output must be used sparingly to avoid it becoming a nuisance. Many warnings would be given only once. Height calls or warnings, if the error condition persisted, were only repeated every 7 s.

(7) The use of speech output to act as a feedback of the DVI recognition process has been demonstrated.

This was not distracting when correctly used and reduced the need to monitor the CDU during data insertion. Better look out resulted. In busy audio environments the pilot filters the relevant information and thus acts upon the most important data. For example, a height call will be ignored when an incoming radio message is present, whereas if a pull up command is heard the radio will be ignored.

(8) Consideration should be given to varying the volume for different warnings such that flight critical messages cannot be missed or mis-heard.

The Wessex trials have clearly indicated that speech recognition and output can give significant benefits to operational effectiveness and improve safety. However, to do this the systems need to be carefully integrated with the many other controls and facilities which make up a total display and flight management system.

Appendix A

THE WESSEX FACILITY

A.1 The cockpit

Fig 1 is a photograph of the Wessex cockpit. The left hand side (LHS) of the instrument panel was the subject pilot's position and the controls and displays available to him were as follows:

Two colour displays (one with a touch overlay)
A control and display unit (CDU) with touch sensitive
Overlay
16 key keyboard
ASR
SSO
Cursor controls and ASR activate switch on cyclic
Map joystick on collective

The right hand side (RHS) instrument panel was used by the observer or safety pilot. Either of the two current display formats could be selected and shown on the RHS tube. A full set of standard electromechanical instruments were retained for use by the safety pilot, for critical sorties such as evaluations of instrument flight in IMC. All the aircrew could hear the SSO.

A.2 Colour displays and formats

The colour displays were standard commercial monitors which were fed with video (domestic standard PAL) signals. A video switching and distribution system enabled any colour CRT to display either of the two formats which were available at any one time. One of these formats which was always displayed, was the primary flight display (PFD). This format displayed sufficient information to enable the pilot to control, navigate, approach and land the helicopter, in visual and instrument flight. Using, dedicated keys or the control and display unit (CDU) the subject pilot could choose to display one of a number of formats on the second multifunction display (MFD) CRT. In addition, with a map format selected the GEC Avionics tactile overlay could be used to designate a waypoint.

A.3 Control and display unit

The control and display unit consisted of a monochrome CRT with a touch sensitive overlay, dedicated keys to the left and right of the CRT, and a numeric keyboard.

The CDU display format was 'menu driven'. By touching the appropriate box on the display the pilot could work his way through the menu. Careful integration of the CDU menu with dedicated keys meant that no more than two selections were required before any data could be inserted or altered.

The basic MENU page is shown in Fig 2. Selection of each option would provide a new set of choices. For example, selection of CHECKS/LIMITS would then provide the pilot with a choice of checks and limits. Some functions produced a further level of

selections. For example if 'navigation' were selected via the dedicated key, followed by 'update' on the touch sensitive overlay, a page would appear asking for the waypoint number and its grid. Further selection of the 'waypoint' dedicated key would put the waypoint listing on the MFD so that as waypoints were updated they could be checked simultaneously for correct entry.

Although most of the CDU menu structure could be accessed quickly, the dedicated keys made it more responsive. These keys short circuited the menu structure and allowed the pilot select directly without starting from the menu page.

A.4 Avionics and interfacing

The flight management processors were interfaced to each other, to the waveform generators and the following systems and sensors:

- Doppler ground speed sensor
- Airspeed
- Heading, pitch, roll attitudes
- Barometric pressure
- Outside air temperature
- Radio altimeter
- Normal and lateral accelerometers
- Collective and cyclic control positions
- Guidance systems — *ILS, MLS and MADGE*.
- Engine and transmission:
  - Power turbine speeds
  - Compressor speeds
  - Power Turbine Inlet Temperatures (PTIT)
  - Fuel flows
  - Fuel contents
  - Rotor rev/min
  - Torque.
- UHF/VHF PTR 1751 radio (digitally controllable)
- Digital map
- Automatic speech recogniser (ASR)
- Synthetic speech output (SSO)
- Non-volatile memory
- Radio clock.

In addition, audio, video and digital recording systems were used to record the display formats, pilot intercom, radio, and the sensor information.

Appendix B

SPEECH RECOGNITION AND ITS IMPLEMENTATION IN THE WESSEX

In recent years many commercial ASR systems have become available which showed promise for use in airborne applications. The ASR system selected for the Wessex trials was an SR-128 manufactured by Marconi Secure Radio Systems. This was a stand-alone device suitable for mounting in a 19 inch rack. It was powered from the 115 V 400 Hz instrumentation inverter to allow it to be used in the helicopter from start-up to shutdown. Speech input was via a 600 ohm balanced line. Recognition and control data were passed on RS 232 serial lines. A photograph of the equipment is shown in Fig 3.

The main features of the SR-128 were as follows:

Speaker trained.
Connected word input
Vocabulary size of 240 words maximum
Memory for 128 s of speech utterances
Syntax programming
Ambient noise compensation (noise mask)
RS 232 interfaces
Self contained mini-cassette handler for program and voice template loading.

B.1 Training

The SR-128 used an acoustic pattern matching technique to recognise words or phrases. This was based upon the principle that for a single speaker the acoustic signature of a particular word would be similar to a pre-recorded version of the same word (the voice template). The user was therefore required to 'train' the ASR by recording the intended recognition vocabulary. Thus the SR-128 can be classified as a speaker dependent system, with any intending user preparing voice templates for subsequent use of the equipment.

For the Wessex trials the templates were prepared in the helicopter, on the ground, without engines running. The pilot used the headset and microphone he intended to use in flight. This ensured that the recorded templates were representative of flight. The headset gave the pilot the audio feedback which he would receive in flight. This was an essential element of the training since varying the feedback volume tended to modulate the power of the users utterances. Once training was complete the templates were stored on tape. Prior to flight, the ASR programme plus the voice template were loaded. This took 2-3 min.

B.2 Connected speech recognition

Current ASR systems can be categorised by three methods of operation: isolated, connected, and continuous word recognition.

Isolated word recognisers can only match single words one at a time, with periods of silence between each word. The user has to speak in an unnatural staccato manner.

Connected word recognisers are able to match words or phrases naturally spoken. However, the recognition process only commences when a period of silence is detected, ie the speaker stops talking. There is therefore a time delay between the speech and recognition occurring. This increases with longer strings of words.

Continuous speech recognisers are able to match words or phrases as they are spoken without waiting for the speaker to pause.

Connected and continuous speech recognition is considerably more difficult than isolated word recognition because:

(a) The recogniser has to choose from a much larger range of options.

(b) Co-articulation will introduce variations in individual words.

and (c) The sound of a word will vary depending upon its prosodic status in a phrase or sentence.

The SR-128 is a connected word recogniser capable of matching a burst of speech of up to 8 s in duration.

B.3 Vocabulary

The maximum number of words or phrases that the SR-128 can handle is 240. Table 1 shows the latest set of words used in the Wessex system. It should be noted that together with the vocabulary size specification, two other parameters were critical. The first was the maximum length of time allowed for the voice templates (128 s) and the second was the maximum duration of an individual word or phrase (2 s for this equipment).

B.4 Syntax

Acoustic pattern matching is a very powerful technique with limited vocabulary sizes (eg 20-40 words). However, as the size of the vocabulary increases there is a greater probability of words or phrases being confused and errors occurring. In addition, acoustic pattern matching of larger vocabularies takes longer. This is particularly so for connected and continuous word recognisers which have to cope with many permutations of word strings. Under such conditions, recognition performance will depend heavily on how close the speakers normal utterances resemble his recorded templates.

To alleviate the problem of vocabulary size the ASR could be programmed to restrict the number of active words at any one time by applying grammatical rules known as syntax. Keywords provided access to different branches of the syntax tree. For example, the keyword 'checks' would enable the work 'take-off' to be activated. The syntax tree currently used in the Wessex is shown in Table 2.

B.5 Noise compensation

In noise environments such as the helicopter, low energy speech features could be completely masked by noise. In addition, the noise presence could significantly effect the way the pilot speaks. As the recognition process depended entirely on pattern

matching techniques, background noise could significantly affect recognition performance.

The SR-128 used a noise mask technique to reduce the effect of the background noise. The ASR was given a noise sample via the pilot's open boom microphone. The power spectrum of the noise was detected by a 19 channel filter of the input to the recogniser. The output of the filter was called the noise mask. With the ASR in recognise mode the noise mask was used in the recognition algorithms to compensate for the noise superimposed on the pilots speech.

Thus there were three distinct actions to complete before the speech recogniser was ready for use. Firstly load the programme, next load the operator's voice template and lastly load the noise mask.

B.6 Interfaces

As words or phrases were recognised the SR-128 provided a stream of data on an RS 232-C serial interface. This stream of data consisted of a start of data code, a three digit template code number, and an end of data code. ASCII characters representing the actual word or phrase and a template score code were also transmitted.

A similar interface was used to prepare the vocabulary and syntax. This was also used to control the operation of the ASR from the cockpit in flight, using the 'experimental' section of the CDU.

B.7 Pilot control

Speech input from the pilots microphone was buffered at the ASR. A rocker switch on the pilots cyclic control column (previously used as the intercom override) was used to activate the pre-amplifier and ASR. This prevented the DVI system from attempting to recognise all of the pilots speech as commands, and to receive only those words intended for it.

Appendix C

TRIALS PROGRAMME AND PROCEDURES

Once the cockpit had been commissioned and the major hardware and software bugs had been ironed out, a period of intense flying was conducted to refine and optimise the use of the system. The aim was to ease the pilot's workload and hence improve mission effectiveness.

Although individual equipment performance was an important factor and, where possible, was measured, emphasis was placed on discovering how and where the new technology could be applied to derive the maximum benefit.

C.1 Pilots

Three Service pilots were used to develop evaluate the total display and flight management system. This enabled pilot opinion expertise, and a knowledge of operational problems from a wide background to be input into the trials. As each pilot had to attain and maintain a good level of understanding and proficiency with the system, it was thought that further pilots would have diluted the amount of useful flying that could be achieved. It should be noted however, that over 80 pilots from the Services, Civil operations and UK Industry have flown in the Wessex and received demonstrations of the equipment and ideas in operation. Their comments have been taken into account during system development and evaluation.

From the initial conception of the programme it was appreciated that the flight management and display system could have an impact on both military and civil operations. Hence tasks were given to the pilots to cover both types of operations, such as general manoeuvres in visual flight, however, others such as nap of the earth (NOE) flight (low level flight using the maximum cover afforded by natural and man-made obstructions), were obviously aimed at the Services. The list below gives an indication of the tasks used and the main reasons for adopting them:

(a) Low level navigation by day in VMC

- navigation accuracy during NOE flight
- hands on control of cockpit management system
- benefits of speech recognition
- benefits of synthetic speech
- benefits of map presentations
- cable/threat warnings display logic
- use of the chin-up CDU during a primarily head-out operation
- overall system effectiveness

(b) Low level navigation at night

- all items listed in (a)
- night vision goggle compatability with digital maps
- primary flight display format development
- overall system assessment during a very demanding task

(c) Medium level navigation in visual and instrument flight

- navigational accuracy compared to task (a)
- performance prediction, fuel, management, endurance monitoring
- radio management in multi radio/frequency areas (*eg* London Control Zone).
- primary flight display development
- multi function display format development
- CDU menu structure

(d) Low speed manoeuvres

- hover display format
- low speed envelope prediction
- digital map targeting
- speech recognition benefits
- synthetic speech benefits

(e) Terminal/approach guidance

- primary flight display development
- synthetic speech during instrument flight
- use of digital maps for self-positioning tasks
- MLS and ILS formats
- speech recognition benefits

(f) General flying/crew training

- development of logic/artificial intelligence to signal malfunctions to the pilot.

Table 1

SPEECH RECOGNITION VOCABULARY

ok	menu
0 1 2 3 4 5 6 7 8 9	map up
take off	hover
land	tees and pees
shutdown	waypoint
reaction	tactical
engines	status
rotor and oil	checks
ILS	limits
MLS	guidance
go around	comms
approach heading	navigation
decision height	rubout
cruise height	flight
cruise speed	ppi
glidepath	data
plot	preset 1
base	preset 2
steer	preset 3
marker	preset 4
update	box 1
route select	box 2
route alpha	icing
route bravo	Bedford tower
route charlie	Bedford approach
survey	Bedford radar
guard	

Table 2 - DVI SYNTAX

Switch on

- Menu-(1)
  - Map up / Map down / Hover / Tees & pees / Waypoint / Tactical -(0)
- Status (2)
  - Menu—(1)
  - Guidance—(4)
  - Comms—(5)
  - Navigation-(6)
  - Map down / Map up / Hover / T & P / Waypoints / Tactical / Flight / PPI / Data / 0-9 / OK / Rub out / Status
- Checks limits (3)
  - Menu—(1)
  - Guidance—(4)
  - Comms—(5)
  - Navigation-(6)
  - Map up / Map down / Hover / T & P / Waypoints / Tactical / Checks / Take off / Land / Shutdown / Reaction / Rotor & oil / Engines / Icing / Limits -(0)
- Guidance (4)
  - Menu—(1)
  - Comms-(5)
  - Nav—(6)
  - Guidance / Map up / Map down / Hover / T & P / Waypoints / Tactical / Approach / Glidepath / Decision Ht / Cruise Ht / Cruise speed / Go around / MLS / ILS / 0-9 / Rub out / Data -(0)
- Comms (5)
  - Menu—(1)
  - Guidance—(4)
  - Navigation-(6)
  - Comms / Map up / Map down / Hover / T & P / Waypoints / Tactical / Preset 1 / Preset 2 / Preset 3 / Preset 4 / Bedford APP / Bedford T/ / Bedford radar / Bedford T/D / Guard / Box 1 / Box 2 / 0-9 / OK / Rub out -(0)
- Navigation (6)
  - Menu—(1)
  - Guidance-(4)
  - Comms—(5)
  - Navigation / Map up / Map down / Hover / T & P / Waypoints / Tactical / Plot / Base / Steer / Marker / 0-9 / Rub out / OK -(0)
  - Route select-(7)
  - Update—(8)
  - Survey—(9)

(7)
- Menu—(1)
- Guidance-(4)
- Comms—(5)
- Navigation / Exit -(6)
- Map up / Map down / Hover / T & P / Waypoints / Tactical / 0-9 / Rub out / OK / Route Alpha / Route Bravo / Route Charlie -(0)

(8)
- Menu—(1)
- Guidance-(4)
- Comms—(5)
- Navigation / Exit -(6)
- Map up / Map down / Hover / T & P / Waypoints / Tactical / 0-9 / Rub out / OK -(0)

(9)
- Menu—(1)
- Guidance-(4)
- Comms—(5)
- Navigation / Exit -(6)
- Map up / Map down / Hover / T & P / WP / Tactical -(0)

(X) Denotes node X

If X = 0 then stays at current node

Table 3

TOTAL SPEECH OUTPUT VOCABULARY

torque	pull up	good morning	good evening
hello	goodbye	activated	ok
ten	twenty	thirty	forty
fifty	sixty	seventy	eighty
ninety	one hundred	one two five	one fifty
two hundred	two fifty	1 hundred & 25	1 hundred & 50
2 hundred & 50	five hundred	thousand	hundred
one	two	three	four
five	six	seven	eight
nine	zero	low	high
deceleration	decision height	glidepath	radio
barometric	look up	height	range
check port T4	checks complete	check stbd T4	Nr high
Nr low	warning	check port compressor	engines
alert	attention	check stbd compressor	attention
check fuel	fuel mismatch	T4 mismatch	Ng mismatch
low fuel	check Nr	reduce power	reduce speed
emergency	check	navigation	comms
guidance	plot		

Table 4

PRIORITY SPEECH OUTPUT VOCABULARY

Event	Speech output
power turbine mismatch > (needle split)	warning engines
stbd engine temperature > limit	check stbd T4
port compressor rpm > limit	check port compressor
stbd compressor rpm > limit	check stbd compressor
fuel flow mismatch > limit	warning fuel mismatch
fuel tank contents mismatch > limit	check fuel
fuel in either tank < limit	warning check fuel
rotor speed out of limits	warning check Nr
torque > limit	torque
cable detected	warning
radar detected	warning
airspeed > maximum allowed (Vmax)	reduce speed
engine temperature mismatch	warning T4 mismatch
compressor mismatch	warning compressor mismatch

IH FS(B) 637

Table 5

OPTIONAL SPEECH OUTPUT VOCABULARY

Event	Speech output
if height < 252 and height > 248	two fifty
if height < 202 and height > 198	two hundred
if height < 152 and height > 148	one fifty
if height < 102 and height > 98	one hundred
if height < 92 and height > 88	ninety
if height < 82 and height > 78	eighty
if height < 72 and height > 68	seventy
if height < 62 and height > 58	sixty
if height < 52 and height > 48	fifty
if height < 42 and height > 38	forty
if height < 32 and height > 28	thirty
if height < decision height	decision height
if height > desired height on directors	high
if height < desired height on directors	low
if height < height for radalt	pull up pull up

TM FS(B) 637

REFERENCES

No.	Author	Title, etc
1	R. Little	The development of an electronic cockpit display and flight management system for helicopters. Technical Memorandum FS(B) 586 (1985)

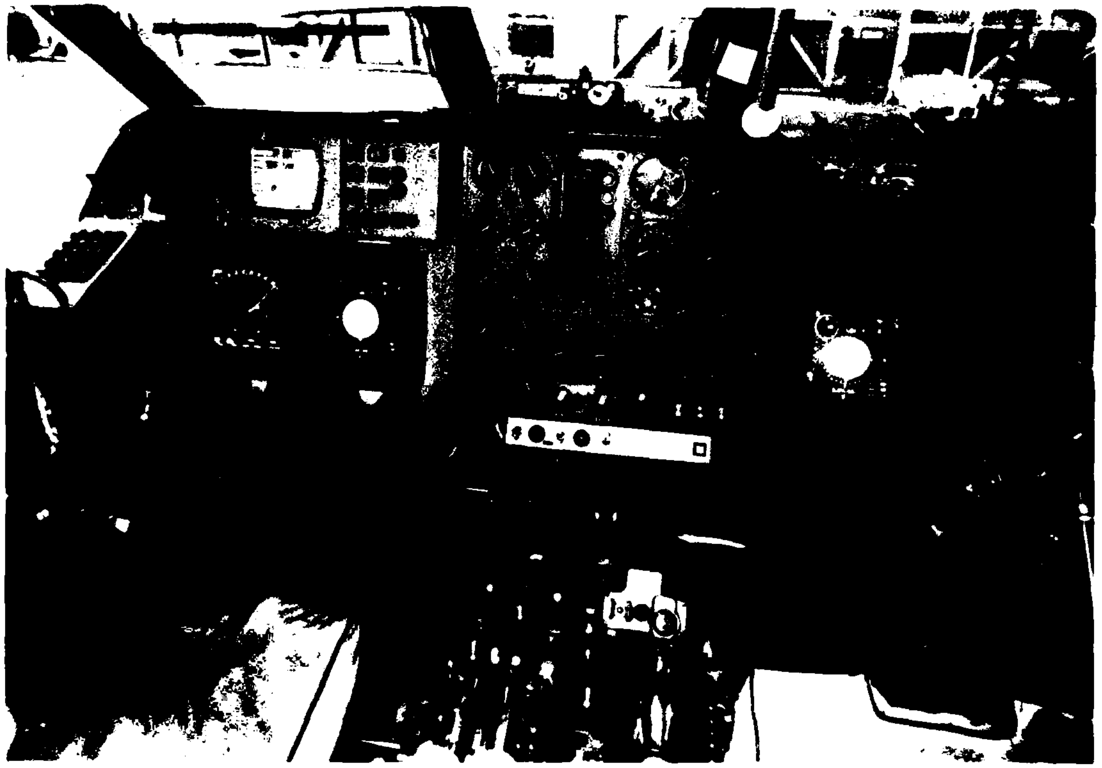


Fig 1 The Wessex cockpit

Fig 2

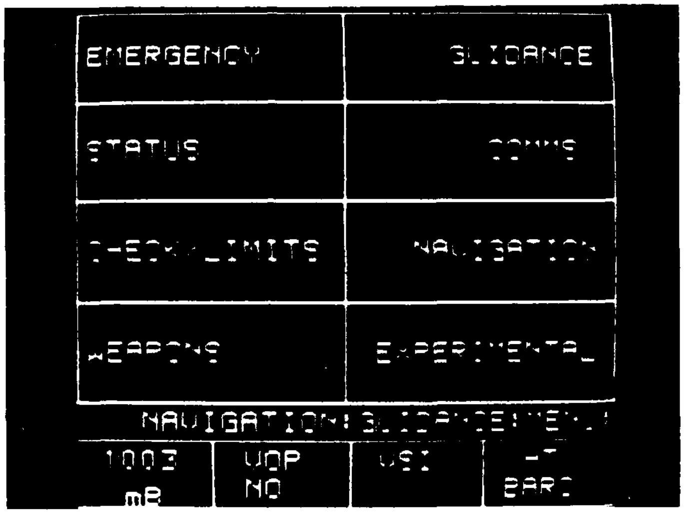


Fig 2 CDU menu page

Fig 3 The SR-128 automatic speech recogniser

Fig 4

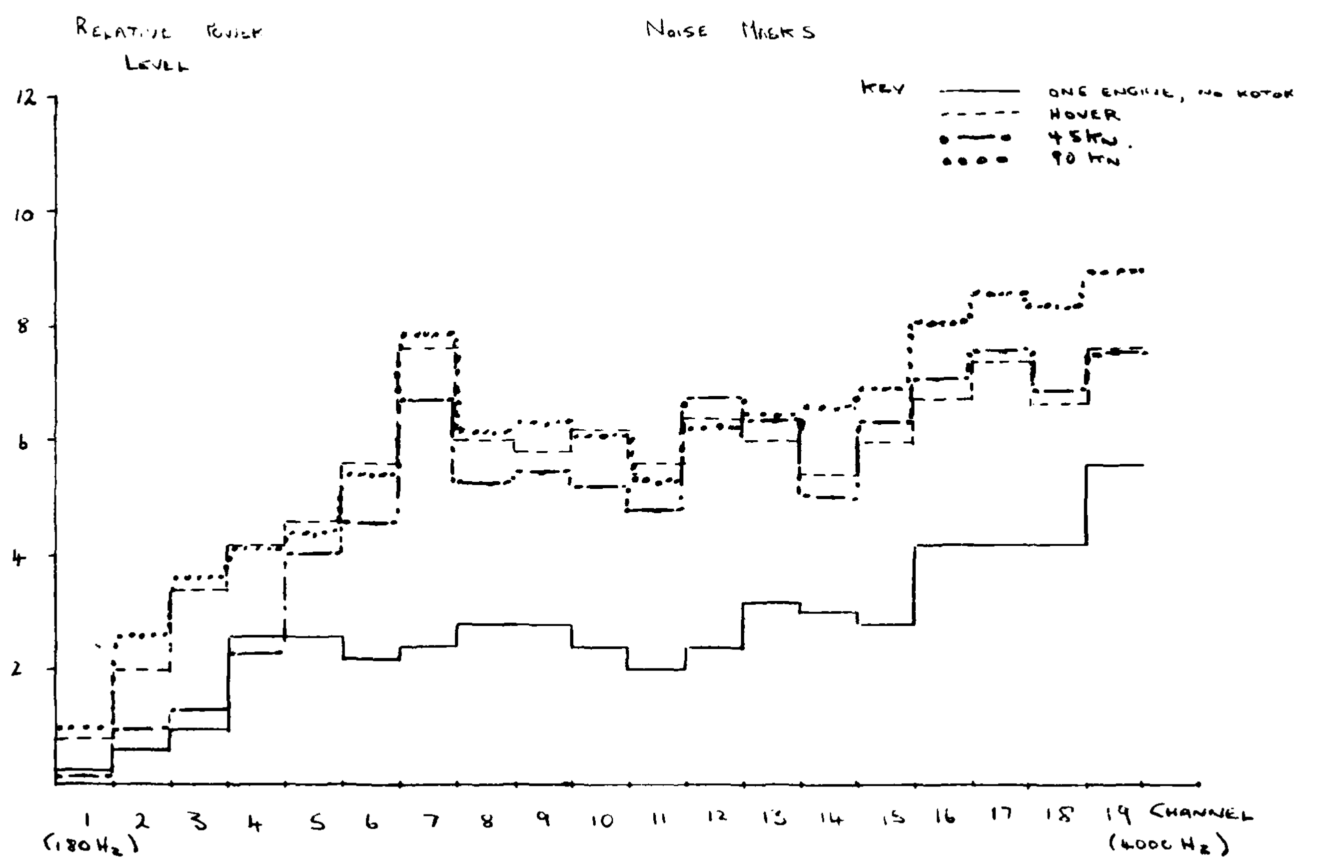


Fig 4 Noise masks in different flight modes (mean values)

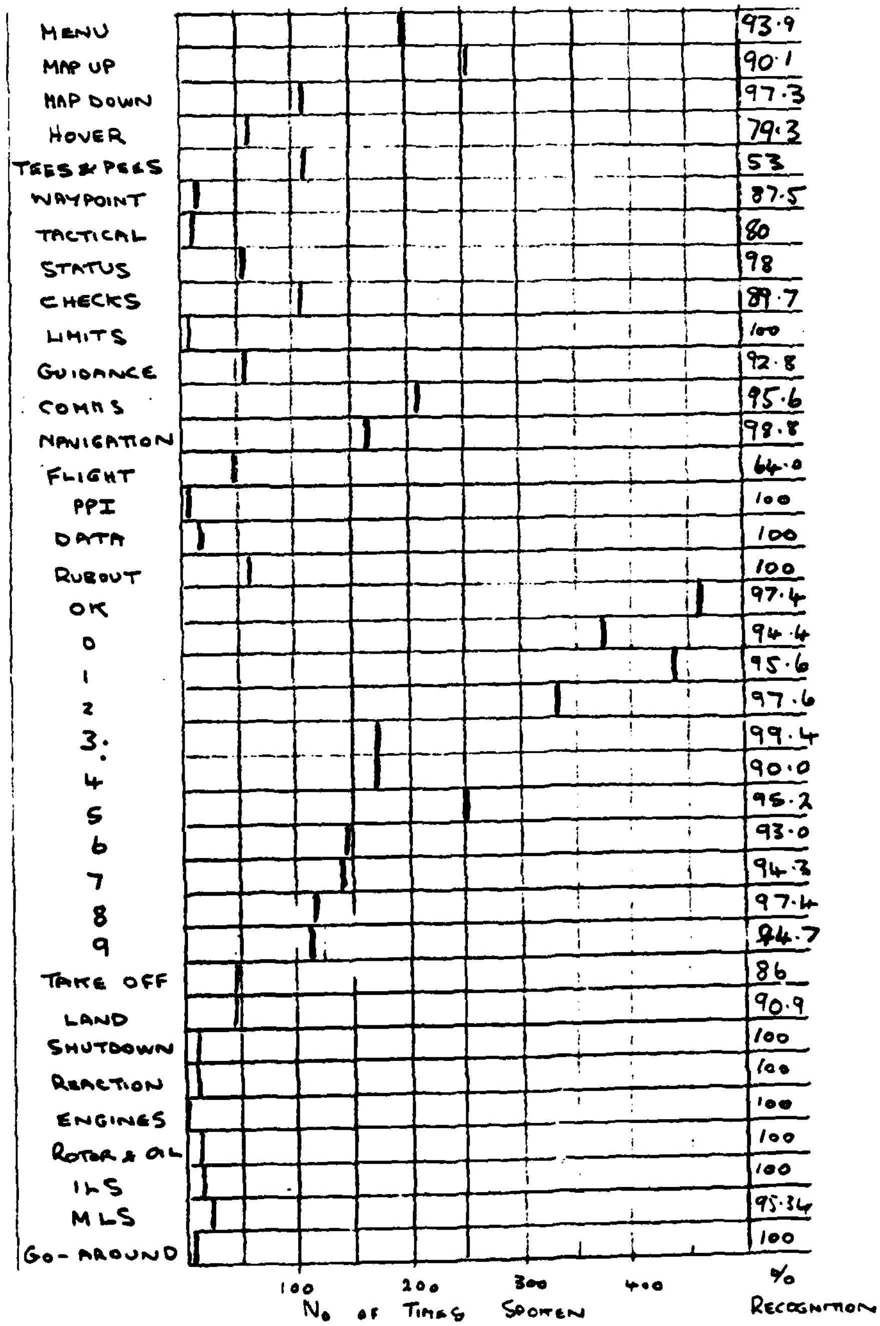


Fig 5 Word use and recognition rates

Fig 6

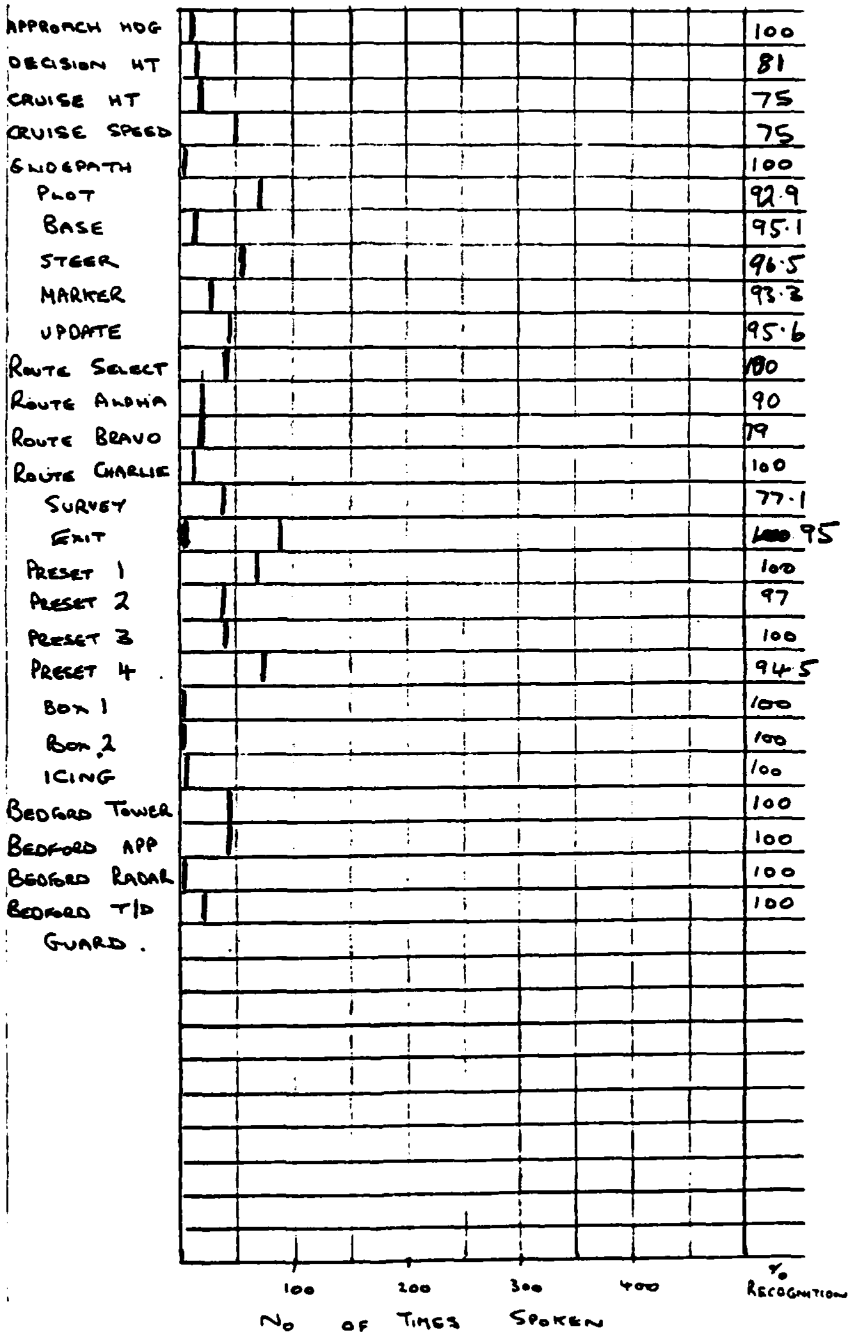


Fig 6 Word use and recognition rates

TM FS(B) 637

OV I RESPONSE

SPOKEN WORD	0	1	2	3	4	5	6	7	8	9	GUIDANCE	MAP UP	COMMS	BASE	OK	EXIT
0	356	–	6	2	–	–		2	–	–	2	–	–	–	–	1
1	–	420	–	–	–	7		2	–	4	–	–	–	1	–	–
2	–	–	323	1	–	–	4	–	–	–	–	–	–	–	2	–
3	–	–	–	167	–	–	1	–	–	–	–	–	–	–		–
4	6	4	2	–	150	–	–	–	–	–	–	–	4	–	–	
5	2	–	1	–	–	237	–	1	–	4	2	–	–	–	–	–
6	–	–	2	–	–	–	133	–	7	–	–	–	–	–	–	–
7	–	–	–	–	–	–	–	134	–	–	–	4	–	–	2	1
8	–	–	–	–	–	–	–	–	114	–	–	–	–	1	–	–
9	–	–	–	–	–	5	–	1	–	108	2	–	–	–	–	1

Fig 7 Confusion matrix for digits

Fig 8

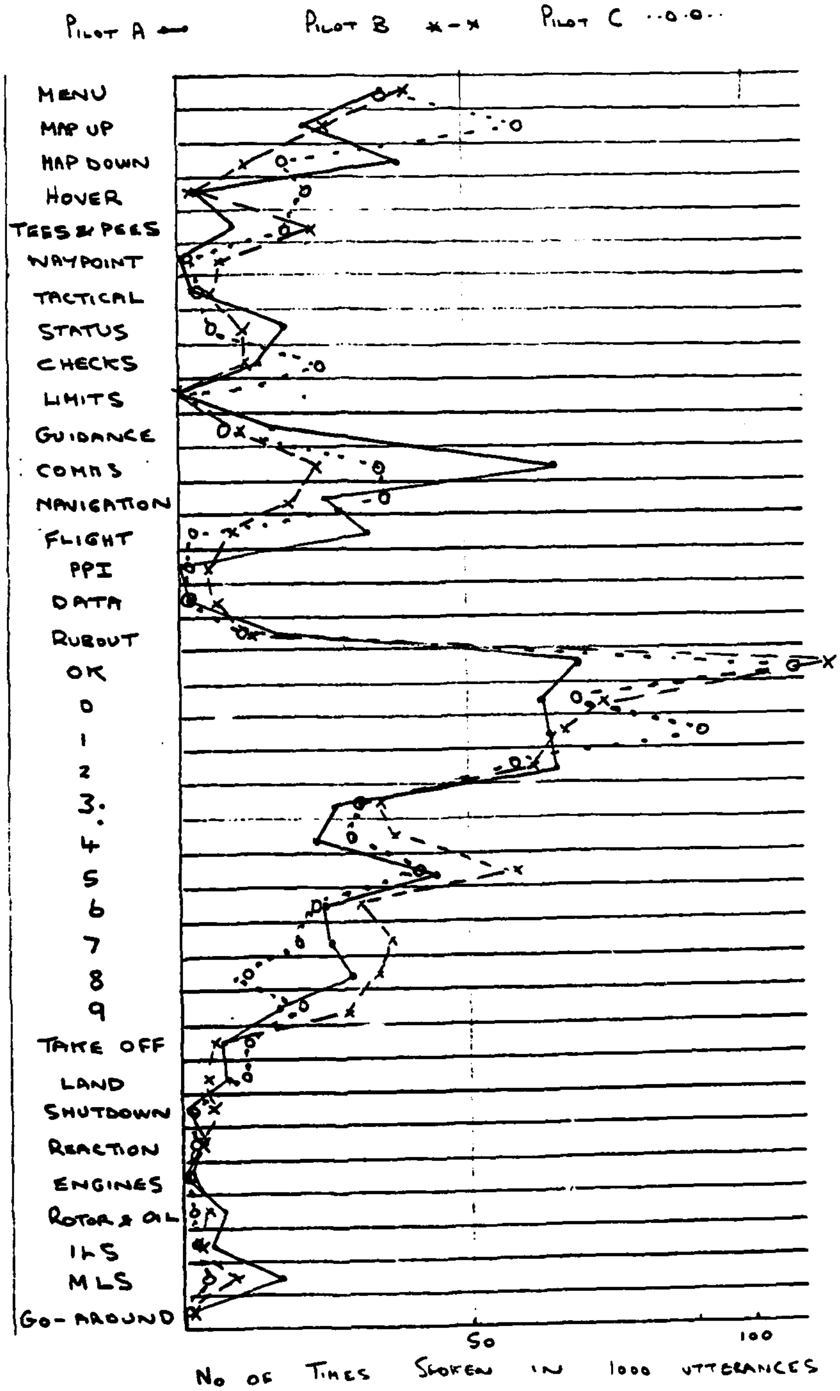


Fig 8 Word usage on DVI

TM FS(B) 637

Fig 9

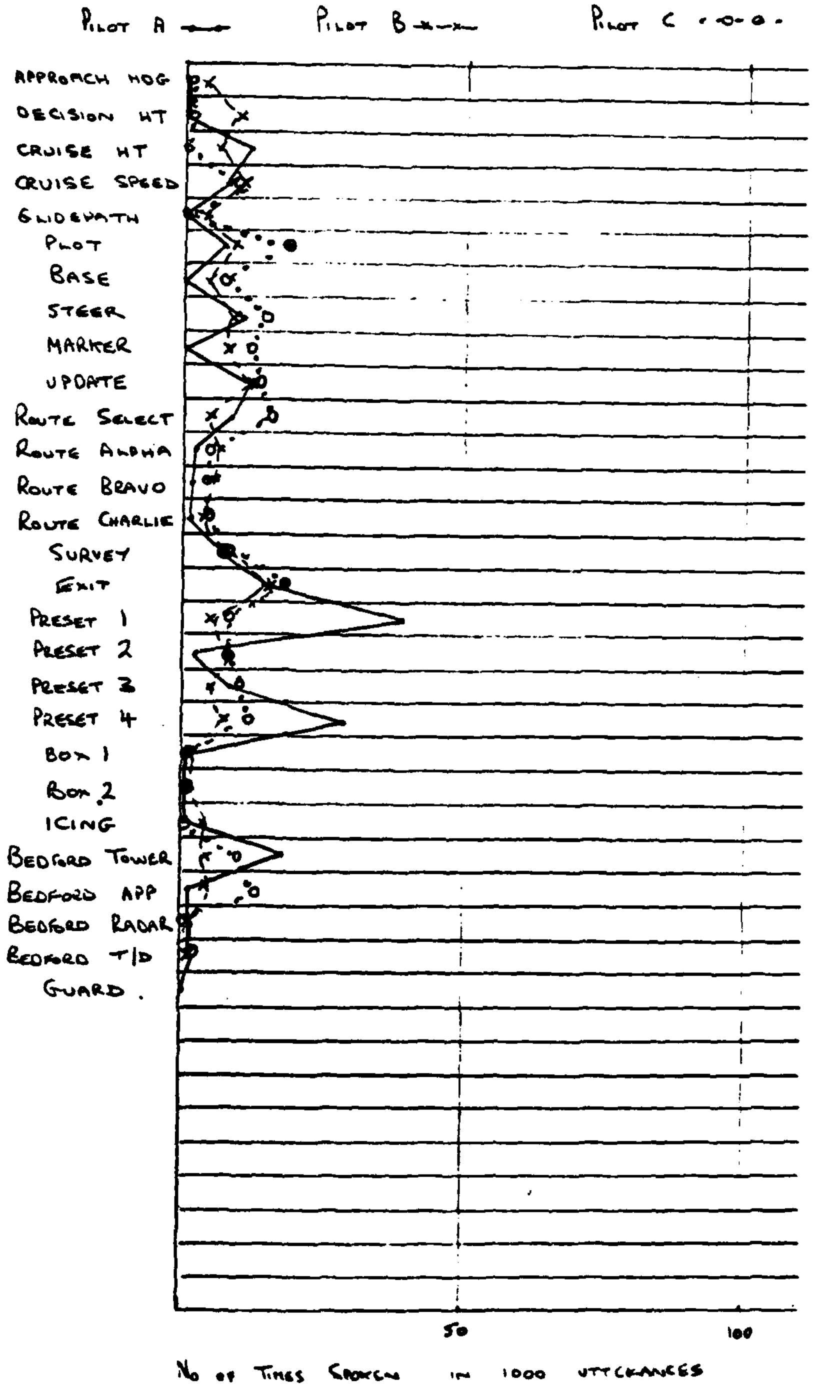


Fig 9 Word usage on DVI

TM FS(B) 637

Fig 10

Response (columns) / **Spoken** (rows)

Spoken \ Response	0	1	2	3	4	5	6	7	8	9
0	356	—	6	2	—	—	—	2	—	—
1	—	420	—	—	—	7	—	2	—	4
2	—	—	323	1	—	—	—	—	—	—
3	—	—	—	167	—	—	1	—	—	—
4	6	4	2	—	150	—	—	—	—	—
5	2	—	1	—	—	237	—	1	—	4
6	—	2	—	—	—	—	133	—	7	—
7	—	—	—	—	—	—	—	134	—	—
8	—	—	—	—	—	—	—	—	114	—
9	—	—	—	—	—	5	—	1	—	108

Fig 10 A confusion matrix for the digits 0-9

TM FS(B) 637

[illegible]

[illegible] If it is necessary to enter classified information, the box [illegible]

1. [illegible]	2. [illegible]	3. Agency Reference	4. Report Security Classification/Marking
[illegible]	[illegible]		UNCLASSIFIED

5. [illegible]	6. [illegible] (Corporate Author) Name and Location
[illegible]	Royal Aircraft Establishment, Bedford, UK
5a. [illegible] Code	6a. Sponsoring Agency (Contract Authority) Name and Location

7. [illegible] [illegible] evaluation of voice interaction as a component of an integrated [illegible] system

7a. [illegible] Title in Foreign Language

7b. [illegible] Title, Place and Date of Conference

8. Author 1. [illegible]	9a. Author 2	9b. Authors 3, 4	10. Date	Pages	Refs.
[illegible]	[illegible], R. Le Cdr		April 1986	42	1
11. [illegible]	12. [illegible]	13. Project	14. Other Reference Nos.		

15. [illegible]

[illegible]

[illegible] RAE Leaflet No.3 to Supplement 6 of [illegible]

[illegible]

END

DATE
FILMED

1 8?